STATISTICAL METHODOLOGY - ASM

DESCRIPTION OF THE ASM SAMPLE

The ASM sample is selected at 5-year intervals beginning the second year subsequent to the Census of Manufactures. The current ASM sample was introduced for survey year 1999 and will be used though the 2003 survey.

In the 1997 Census, there were approximately 365,000 individual manufacturing establishments. For sample efficiency and cost considerations, the 1997 manufacturing population was partitioned into two components: a mail stratum and a nonmail stratum.

Mail Stratum

The mail stratum of the survey is comprised of larger single-location manufacturing companies and all manufacturing establishments of multi-unit companies (companies that operate at more than one physical location). Approximately 200,000 of the 365,000 establishments in the 1997 census were assigned to the mail stratum.

On an annual basis, the mail stratum is supplemented with larger, newly active single-location companies identified from a list provided by the Internal Revenue Service (IRS) and new manufacturing locations of multi-unit companies identified from the Census Bureau's Company Organization Survey (COS).

For the 1999 survey, a new sample of approximately 52,000 individual establishments was selected from the mail stratum assembled from the 1997 census. Supplemental samples representing both 1998 and 1999 births (newly active establishments that were not included in the 1997 Census) were also selected. Establishments selected for the sample will be included in the ASM through 2003.

The 1999-2003 ASM sample design is similar to the one that has been used since 1984; however, several adjustments to the arbitrary certainty portion of the sample were made. Companies in the 1997 Census with manufacturing shipments of at least $1.0 billion were defined as company certainties. The threshold for earlier surveys was $500 million. For these large companies, each manufacturing establishment is included in the ASM mail sample with certainty. For the 1999-2003 sample, there are approximately 500 certainty companies collectively accounting for over 14,000 establishments. This is a reduction from the 1994-98 sample where there were 650 certainty companies accounting for over 18,000 establishments.

For the remaining portion of the mail component of the survey, the establishment was defined as the sampling unit. All establishments with 500 employees or more were defined as employment certainties. (This threshold was raised from 250 in the 1994-98 sample.) In addition, all establishments producing products in NAICS industry 334111, Electronic Computer Manufacturing, were defined as certainties. A final certainty category, small industry certainties, was introduced in 1999. This was designed to

eliminate estimation difficulties for industries with few establishments. For 6-digit NAICS industries that contained 20 or less establishments in the mail stratum, all establishments were included in the sample with certainty. Across these four arbitrary certainty classes, there were approximately 16,600 establishments. Collectively, they accounted for approximately 62 percent of the total value of shipments in the 1997 Census of Manufactures.

Smaller establishments in the remaining portion of the mail stratum were sampled with probabilities ranging from .02 to 1.00. The initial probabilities of selection assigned to these establishments were proportionate to a measure-of-size determined for each establishment. The measure-of-size was a function of the establishment's 1997 industry classification and its 1997 product class data. For each product class and industry, a desired reliability constraint was specified. Using a technique developed by Dr. James R. Chromy of the Research Triangle Institute, the initial establishment probabilities were optimized such that the expected sample simultaneously satisfied all industry and product class reliability constraints while minimizing the sample size. This technique reduces the likelihood of selecting nonrepresentative samples for individual product classes or industries.

This method of assigning probabilities based on product class shipments is motivated by our primary desire to produce reliable estimates of both product class and industry shipments. The high correlation between shipments and employment, value-added, and other general statistics assures that these variables will also be well represented by the sample. The actual sample selection process uses an independent chance of selection method (Poisson sampling).

Nonmail stratum

The initial nonmail component of the survey was comprised of approximately 161,500 small, single-establishment companies in the 1997 Census of Manufactures; and approximately 3500 relatively small single-location companies that had been selected for the 1994-98 ASM sample. The addition of this second group to the nonmail stratum prevented these companies from been selected for two consecutive ASM samples. The initial nonmail stratum was comprised of approximately 165,000 establishments. The nonmail stratum is supplemented annually using the list of newly active single-location companies provided by the IRS and payroll cutoffs. Companies with payroll below the payroll threshold are added to the nonmail stratum.

For this portion of the population, sampling is not used. The data for this group are estimated based on selected information obtained annually from the administrative records of the IRS and Social Security Administration (SSA). This administrative information, which includes payroll, total employment, industry classification, and physical location, is obtained under conditions which safeguard the confidentiality of both tax and census records.

DESCRIPTION OF THE ASM ESTIMATION PROCEDURE

Most of the ASM estimates derived for the mail stratum are computed using a difference estimator. At the establishment level, there is a strong correlation between the current-year data values and the corresponding 1997 (base) data values. Therefore, within the mailed stratum, for each item at each level of aggregation, an estimate of the "difference" between the current year and the base year is computed from sample cases and added to the corresponding base-year values. For the 1998-2001 ASM estimates, the 1997 Census of Manufactures values serve as the base year.

Because of the positive year-to-year correlation, estimates derived using this methodology are generally more reliable than comparable estimates developed from the current sample data alone. Estimates for the capital expenditures variables are not generated using the difference estimator because the year-to-year correlations are considerably weaker. The standard linear estimator is used for these variables.

For the nonmail stratum, estimates for payroll and employment are directly tabulated from the administrative-record data provided by IRS and SSA. Estimates of data other than payroll and employment are developed from industry averages. Although the nonmail stratum contains approximately 200,000 individual establishments in 1999, it accounts for less than 2 percent of the estimate for total value of shipments at the total manufacturing level.

Corresponding estimates for the mail and nonmail components are combined to produce the published estimates.

QUALIFICATIONS OF THE ASM DATA

The estimates developed from the sample are apt to differ somewhat from the results of a survey covering all companies in the sample lists but otherwise conducted under essentially the same conditions as the actual sample survey. The estimates of the magnitude of the sampling errors (the difference between the estimates obtained and the results theoretically obtained from a comparable, complete-coverage survey) are provided by the standard errors of estimates.

The particular sample selected for the ASM is one of many similar probability samples that, by chance, might have been selected under the same specifications.
Each of the possible samples would yield somewhat different sets of results, and the standard errors are measures of the variation of all the possible sample estimates around the theoretically comparable, complete-coverage values.

Estimates of coefficient of variation (expressed as %) and standard errors have been computed for selected ASM statistics in this report where appropriate.

In conjunction with its associated estimate, the standard error or coefficient of variation may be used to define confidence intervals (ranges that would include the comparable, complete-coverage value for specified percentages of all the possible samples). The complete-coverage value would be included in the range:

From one standard error below to one standard error above the derived estimate for about two-thirds of all possible samples.

From two standard errors below to two standard errors above the derived estimate for about 19 out of 20 of all possible samples.

An inference that the comparable, complete-survey result would be within the indicated ranges would be correct in approximately the relative frequencies shown. Those proportions, therefore, may be interpreted as defining the confidence that the estimates from a particular sample would differ from complete-coverage results by as much as one, two, or three standard errors, respectively.

For example, suppose an estimated total is shown at 50,000 with an associated relative standard error of 2 percent, that is, a standard error of 1,000 (2 percent of 50,000). There is approximately 67 percent confidence that the interval 49,000 to 51,000 includes the complete-coverage total, about 95 percent confidence that the interval 48,000 to 52,000 includes the complete-coverage total, and almost certain confidence that the interval 47,000 to 53,000 includes the complete-coverage total.

In addition to the sample errors, the estimates are subject to various response and operational errors: errors of collection, reporting, coding, transcription, imputation for nonresponse, etc. These operational errors also would occur if a complete canvass were to be conducted under the same conditions as the survey. Explicit measures of their effects generally are not available. However, it is believed that most of the important operational errors were detected and corrected during the Census Bureau's review of the data for reasonableness and consistency. The small operational errors usually remain. To some extent, they are compensating in the aggregated totals shown. When important operational errors were detected too late to correct the estimates, the data were suppressed or were specifically qualified in the tables.

As derived, the estimated standard errors included part of the effect of the operational errors. The total errors, which depend upon the joint effect of the sampling and operational errors, are usually of the order of size indicated by the standard error, or moderately higher. However, for particular estimates, the total error may considerably exceed the standard errors shown.

1999 Computer Network Supplement

The 1999 Annual Survey of Manufactures Computer Network Use Supplement was mailed to the plants included in the ASM sample. This supplement collected information about manufacturers= e-commerce activities and e-business processes. The questionnaire asked the plant if they accepted online orders and the percentage of total shipments that were ordered online. Information on online purchases was also requested. In addition, information was collected about the plants= current and planned use of selected e-business processes and the extent to which the plant shared information online with its vendors, customers, and other plants within the company.

Approximately 83 percent of the plants responded to the supplement. A stratified random sample of approximately 150 nonrespondents was selected. These plants were contacted by telephone to determine if they accepted online orders and to obtain the percentage of total shipments ordered online, if appropriate. The information collected from this sample was weighted to represent the entire group of nonrespondents.

Estimates for NAICS subsectors were calculated from the respondents to the supplement by summing the reported data weighted by the inverse of the probability of selection of the establishment=s inclusion in the ASM. Estimates from the supplement and the nonresponse sample were combined to represent the entire ASM sample. These estimates of online shipments are not directly comparable to the estimates of total value of shipments published in the ASM due to differences in the estimator and coverage of the manufacturing sector. Two separate adjustments to the online shipments estimates were made to improve the comparability. These adjustments varied by NAICS subsector and were generally in the range of 3-4 percent.